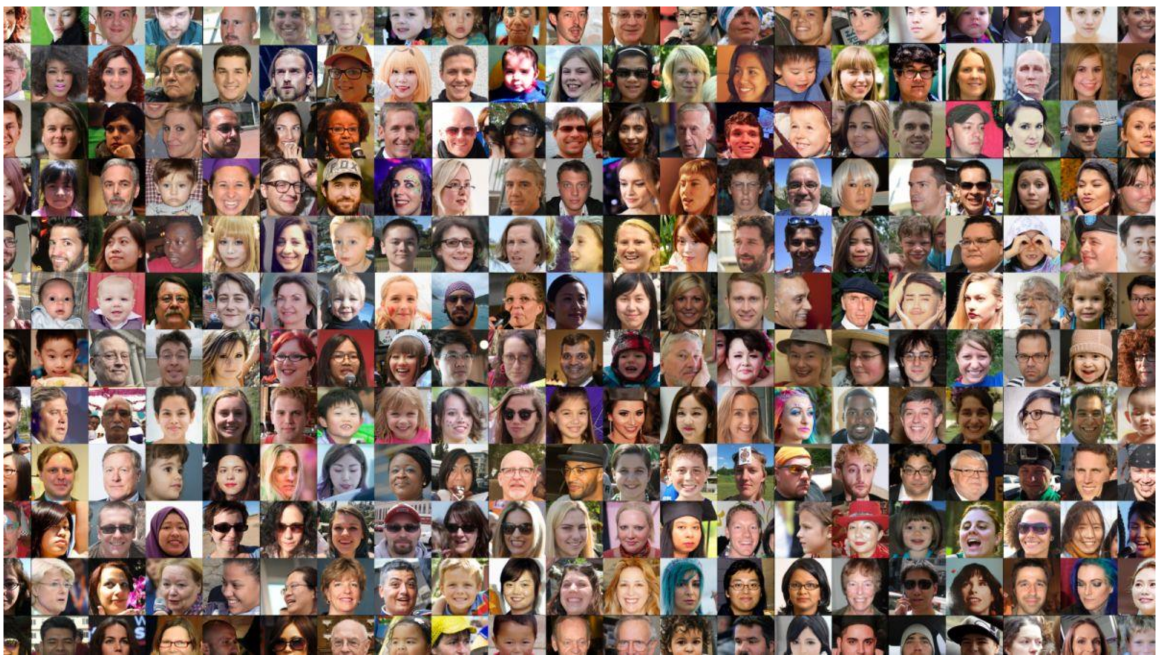


Facial Recognition (Initial Project)

CREATING A FULLY OPERATIOONAL FACIAL RECOGNITION PROGRAM

Aran Miller | Data Science Capstone | 2 June 2021

Abstract

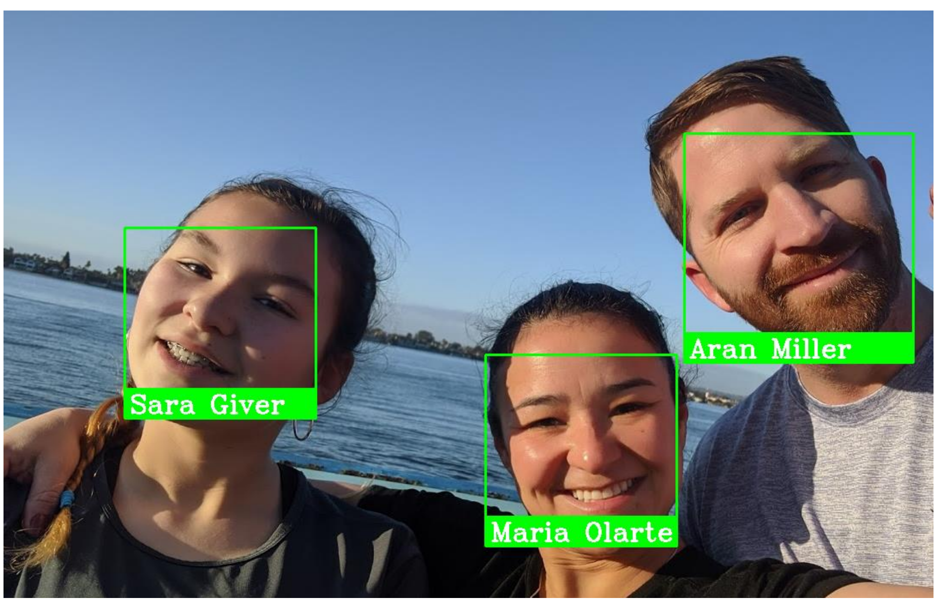
Facial Recognition has been a very hot topic since early 2000s. While most people are similar with Facial Recognition through their phone recognizing them or Facebook and Google auto tagging themselves and their friends and family in photos uploaded to the service, facial recognition has been growing in law enforcement. We have truly gone from TV shows and movies depicting computers matching individuals as science fiction to this technology being used on a daily basis at multiple agencies in government.

Research Question

How is it possible for a computer to understand and then classify and identify faces in photographs? Once the computer can identify what a face is, it can then train and learn known persons faces to then identify unknown persons. The computer will be trained on these unknown persons and then be scored based on how accurately it can identify images of persons who are both included and excluded from the known persons.

Dataset Description

The dataset used in this project was generated using images collected of TV and Movie actors along with a handful of known persons close to this project. The dataset contains 554 images. These images vary in size and quality and are all stored in RGB color. These images are stored as images in many image formats, such as .jpg and .png formats. These images will also be labeled based on the person or persons contained within the image itself.



This data-set was split into a Train and Test set. The training set contains 98 images and accounts for about 17.7% of the data. While this is relatively low for a training set the nature of facial recognition allows the training set to be relatively small and unless more known persons are added to the dataset there are diminishing returns to adding more and more photos of known persons who are already located within the dataset. The Test set is split into two further sets, an Uknown set of unlabeled data to be used during testing and to give a demonstration of the software capabilities and an Accuracy set of 404 images which is used to determine how accurate the model performs.

In these data-sets the images were places into 1 of 11 categories. Ten of these categories were known persons, such as celebrities, and the final label was "unknown" meaning they were photos of people but they were not individually labeled as who was in the photo specifically.

Steps of Facial Recognition

There are five major aspects of facial recognition. These five aspects are face detection (finding a face in an image), generating landmark locations (identifying eyes, nose, lips), transform/warp the image so the landmark locations are in a centralized location so all faces can be compared easily, determine features of images to easily encode images to compare images more easily and faster, finally, comparing two faces and classifying an unknown image based off known images.

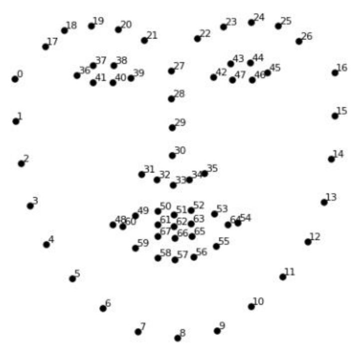
Face Detection

Face detection is the first major aspect of facial recognition. How you can compare two faces if you don't know where the faces are in the image? The first major face detection method that was created was the Caffe model which was quickly replaced by the Haar Cascade method. Haar Cascade was only overtaken in popularity when Histogram of Oriented Gradients. Both Haar Cascade and Histogram of Gradients are both still widely used but as Convolutional Neural Networks become more robust at classifying objects it is becoming more accurate and more widely accepted as the go to standard for face detection.

While all four of these methods and models were used in various aspects of this project, the major focus was Histogram of Gradients because of its efficiency running on a CPU. The HOG method will first take an image and convert it into a black and white image. Once in black and white each pixel will generate an arrow towards one of the eight pixels around it, and this arrow will face the direction of the darkest pixel. These arrows will then generate a separate image that will more accurately show the outline of a face in the image. To generate the model, millions of these images will generate a "template" for what a face looks like and thus it can find faces inside of any image more accurately.

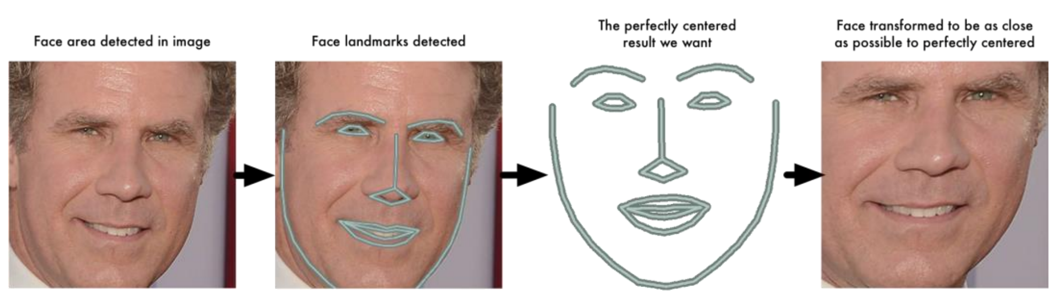
Generating Landmarks

Once the HOG image is generated the arrows will create the landmarks for the face. These landmarks will outline the overall shape of the face around the cheeks and chin as well as eye brows, eyes, nose, upper lip and lower lip. The standard set of landmarks has been accepted to be 68 landmarks.



Warp Image

Once the 68 landmarks are generated for an image the is warped so that the facial features are centered into the image itself. This allows all images to have a similar face shape regardless of the angle or direction the face is in the photo. This obviously has its limitations, such as if someone is completely profile in a photo, many of the landmarks would not be available.



Face Encoding

Now that the faces have been found and standardized, we are able to train a model to learn when to images are of the same or different person. This is more commonly done with a Convolutional Neural Network, specifically a Triplet Loss function is used. A Triplet Loss will take three images, two of the same persons, and one of a different person. The model will then generate a 128 point vector to "identify" a face. These generated vectors will then change based off the following loss function:

$$L(A, P, N) = \max(D(A,P) - D(A,N) + alpha, 0)$$



In this function, A, P, N stand for Anchor, Positive, Negative. The anchor is an image, the Positive is a different image of the same person as the Anchor, and Negative is a separate image, but is a different person than the Anchor. D (X, Y) is the distance between the two vectors. So, as the machine learns, vectors A and P will become closer while A and N become further apart. This method of picking 3 images, two of the same person, and one of a separate person will be processed tens of millions of times, on millions of images, and hundreds of thousands of different identities.

Once the network as fully trained the model it can generate a 128 point vector for any face. This essentially is like the model creating a fingerprint for the face itself. When two of these vectors are compared to each other if they are "close" then they are likely to be the same person.

Face Classification

The final step to facial recognition is generating these 128 vectors for a known list of labeled individuals. This is similar to a training set for the model. Once the array of known person vectors has been generated these are input as the features for any classification model. For this project, both a Support Vector Machine Classifier and a K-Nearest Neighbors Classifier were used to predict identities of an unknown person in a photo. For each unknown image that wants to be tested to identify individuals, the faces are detected in the photo, and then generate a 128 point vector for the image and compared to the fitted model. The prediction will be the closest match against the known list.

Results

While any classifier model can work for this final step, the focus of this project was comparing SVM and KNN models. When testing to determine the accuracy of the models they would actually predict the same results within a very slim margin. Both were tested using different thresholds of when to classify an image as a known persons or an unknown persons. Using the best threshold for both models they both received an overall accuracy of 98.995%. This shows to me that the method and models for training the face detection and more so the face encoding is much more important than the classifier itself when identifying faces.

Conclusion and future goals

Overall this was an incredible project and gave true insight into what it takes to generate a fully fledged face recognition model. While it is widely accepted to use pretrained models for face detection and face encodings, I would really like to create a facial recognition program from the ground up and be able to control these algorithms and how they are implemented.

I will be continuing this project by building a dataset of 40-50 million labeled individuals to train my own face detection model and my own face encoding model. Facial recognition performs much better on western individuals compared to middle eastern, African, or Asian diversities. This bias stems from the datasets that are used to generate these models. The latter of these groups are underrepresented and thus are much harder to identify. My goal is to attempt to eliminate these biases in an effort to make facial recognition programs even more accurate.

My hopes are to generate this program from the ground up over the next year or two of my career. Once this project and goal are completed I would like to further my skills and research into more general CNN models. As these models evolve to be more generalized in object detection this creates a great opportunity for computer vision, such as self-driving cars and applications that haven't even been considered yet.